Virology 266, 8–16 (2000)
doi:10.1006/viro.1999.0044, available online at http://www.idealibrary.com on IDEAL®

RAPID COMMUNICATION

Large-Scale Monitoring of Host Cell Gene Expression during HIV-1 Infection Using cDNA Microarrays

Gary K. Geiss,* Roger E. Bumgarner,* Mahru C. An,* Michael B. Agy,† Angélique B. van 't Wout,* Erick Hammersmark,‡ Victoria S. Carter,† David Upchurch,* James I. Mullins,* and Michael G. Katze*,†,[1]

*Department of Microbiology and ‡Department of Molecular Biotechnology, School of Medicine and †Washington Regional Primate Research Center, University of Washington, Seattle, Washington 98195

Received September 2, 1999; returned to author for revision October 1, 1999; accepted October 12, 1999

Human immunodeficiency virus type 1 (HIV-1) infection alters the expression of host cell genes at both the mRNA and protein levels. To obtain a more comprehensive view of the global effects of HIV infection of CD4-positive T-cells at the mRNA level, we performed cDNA microarray analysis on approximately 1500 cellular cDNAs at 2 and 3 days postinfection (p.i.) with HIV-1. Host cell gene expression changed little at 2 days p.i., but at 3 days p.i. 20 cellular genes were identified as differentially expressed. Genes involved in T-cell signaling, subcellular trafficking, and transcriptional regulation, as well as several uncharacterized genes, were among those whose mRNAs were differentially regulated. These results support the hypothesis that HIV-1 infection alters expression of a broad array of cellular genes and provides a framework for future functional studies on the differentially expressed mRNA products.

Key Words: cDNA; microarray; gene expression; HIV-1; host cell.

Viruses, such as human immunodeficiency virus type 1 (HIV-1), impact on a diverse set of the host cell's biochemical processes. For example, viral infection induces the interferon antiviral response (*1, 2*), modulates expression of cell surface molecules, and manipulates the cell's transcriptional, translational, and trafficking machinery (*3*). Many of these interactions will be characterized by changes in cellular mRNA levels that will depend on both the stage of infection and the biological state of the infected cell. The recent emergence of high-density DNA arrays (microarrays or oligonucleotide chips) has revolutionized gene expression studies by providing a means to measure mRNA levels for thousands of genes simultaneously [for review see (*4–6*)]. The potential of this technology for investigating viral–host cell interactions has yet to be realized, as it has only recently been applied to a viral infection (*7*). Here we used cDNA microarray technology to examine HIV-1-LAI infection of the CD4+ T-cell line CEM-CCRF in order to identify cellular genes influenced by viral infection. In addition, custom software was developed to quantitatively analyze the microarray data. We have identified 20 cellular mRNAs from a variety of cellular pathways, whose levels are altered by HIV-1 infection.

Results. Differential expression of host cell genes in HIV-1-infected CEM cells was investigated with the use of cDNA microarray analysis. Polyadenylated RNA was isolated from mock- and HIV-1-infected cells at 2 and 3 days p.i. Earlier work from our laboratory has demonstrated that by 2 days p.i. approximately 80% of the cells were infected and by 3 days p.i. more than 95% of the cells were infected under the experimental conditions utilized (*8*). Poly(A)$^+$ mRNAs were reverse transcribed in the presence of either Cy3- or Cy5-labeled dCTP to produce fluorescently labeled first-strand cDNAs. The mock-infected sample was labeled with Cy3 and the HIV-1-infected sample was labeled with Cy5. To reduce potential biases in labeling and hybridization due to characteristics of the particular Cy-labeled dCTP, this labeling scheme was reversed in half of the experiments (i.e., mock-infected Cy5 and HIV-1-infected Cy3 cells). The appropriate samples were combined and hybridized to replicate slides. All slides were made with duplicate sets of 1506 target cDNAs on the left- and right-hand sides of the array. The microarrays were scanned with a dual-laser scanning confocal microscope (Molecular Dynamics) producing two images, each corresponding to one-half of the slide. A pseudo-color image of the hybridization results for the entire 1506 gene set in each labeling scheme is shown (Fig. 1A). A magnified view of one section of replicate microarrays hybridized with samples from both 2 (Fig. 1B) and 3 (Fig. 1C) days p.i. is also

[1] To whom correspondence and reprint requests should be addressed at Department of Microbiology, School of Medicine, Box 357330, Seattle, WA 98195. Fax: (206)-685-0305. E-mail: honey@u.washington.edu.

0042-6822/00 $35.00

TABLE 1

List of Differentially Expressed Genes

Accession No.	Putative gene ID	Fold change (2-day PI)	Fold change (3-day PI)
	Upregulated genes		
AA427491	T-cell receptor α-chain, c region	1.19 ± 0.14	3.13 ± 0.77
AA443584	CD8-α precursor	0.99 ± 0.14	1.63 ± 0.24
AA455278	Human neurogranin	1.02 ± 0.07	1.68 ± 0.24
AA458503	Receptor-interacting protein 140 (RIP140)	0.84 ± 0.10	1.74 ± 0.17
AA457123	Transfer valyl-tRNA synthetase	0.89 ± 0.13	1.68 ± 0.14
AA457118	hTOM34 translocase	1.03 ± 0.13	1.82 ± 0.12
AA290698	IMAGE ID 700414	1.21 ± 0.14	2.23 ± 0.27
H47026	IMAGE ID 178468	0.93 ± 0.16	1.52 ± 0.11
AA455695	IMAGE ID 813990	1.10 ± 0.09	1.66 ± 0.25
	Downregulated genes		
AA455931	Adenylosuccinate lyase	1.00 ± 0.21	0.65 ± 0.02
M26708	Prothymosin alpha	0.81 ± 0.14	0.48 ± 0.04
L00160	Phosphoglycerate kinase	1.17 ± 0.21	0.58 ± 0.13
AA401111	Glucose phosphate isomerase	1.04 ± 0.07	0.51 ± 0.04
AA621300	p18 protein mRNA/stathmin	0.95 ± 0.12	0.60 ± 0.10
AA463631	Signal recognition particle 72	0.66 ± 0.16	ND
AA399473	Placental protein 5 (PP5)	ND	0.67 ± 0.09
AA504327	Type IVA protein tyrosine phosphatase	0.83 ± 0.13	0.57 ± 0.06
U34379	Txk kinase	0.65 ± 0.15	ND
AF038953	E25 mRNA (clone 1)	1.21 ± 0.13	0.57 ± 0.04
	E25 mRNA (clone 2)	1.15 ± 0.08	0.61 ± 0.06

Note. ND, not determined due to insufficient signal above background.

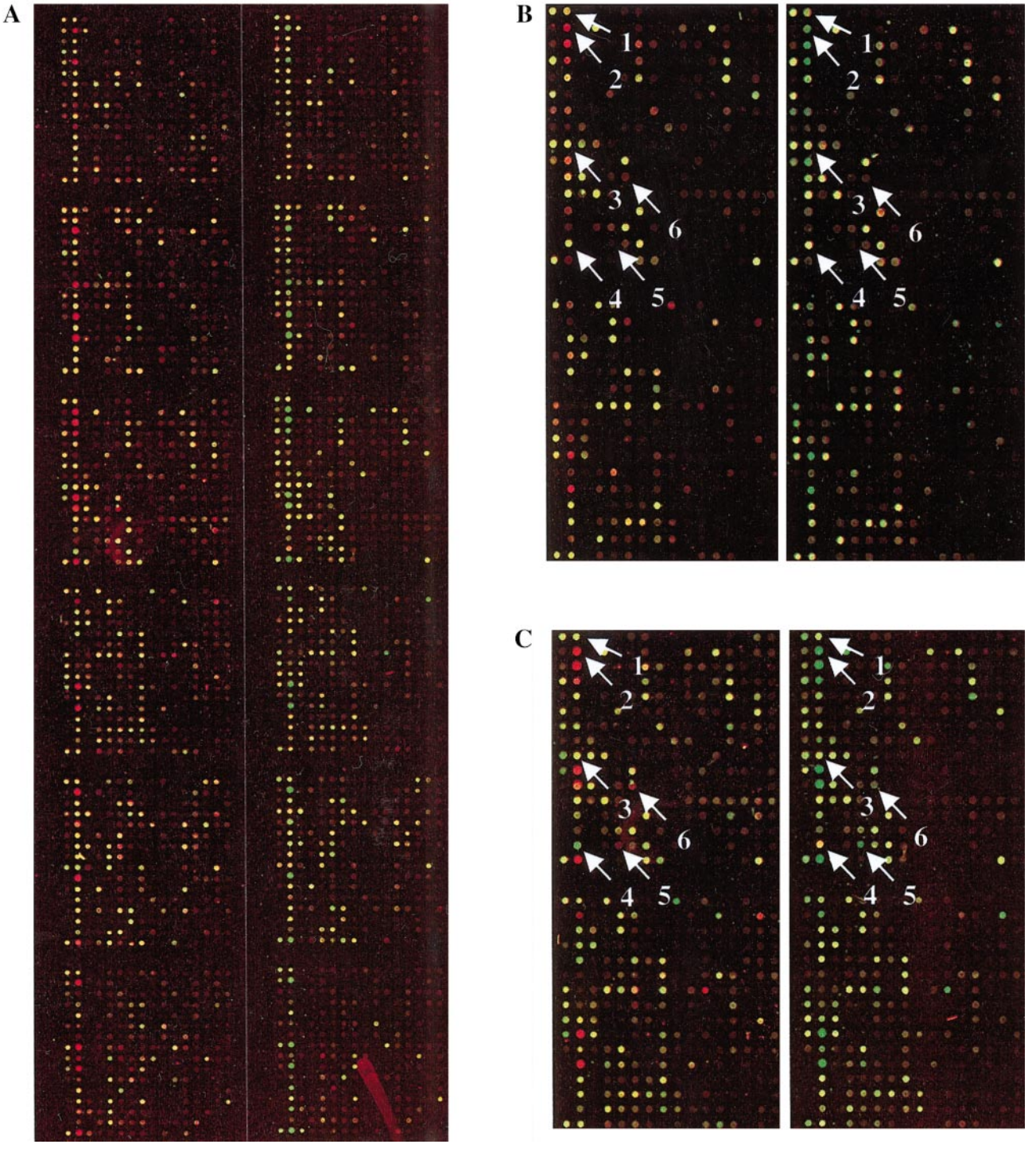
shown. The entire data set consists of four replicate microarrays (two in each color scheme) for 2 days p.i. and two replicate microarrays for 3 days p.i. (one in each color scheme), resulting in eight and four separate measurements per gene for days 2 and 3 p.i., respectively.

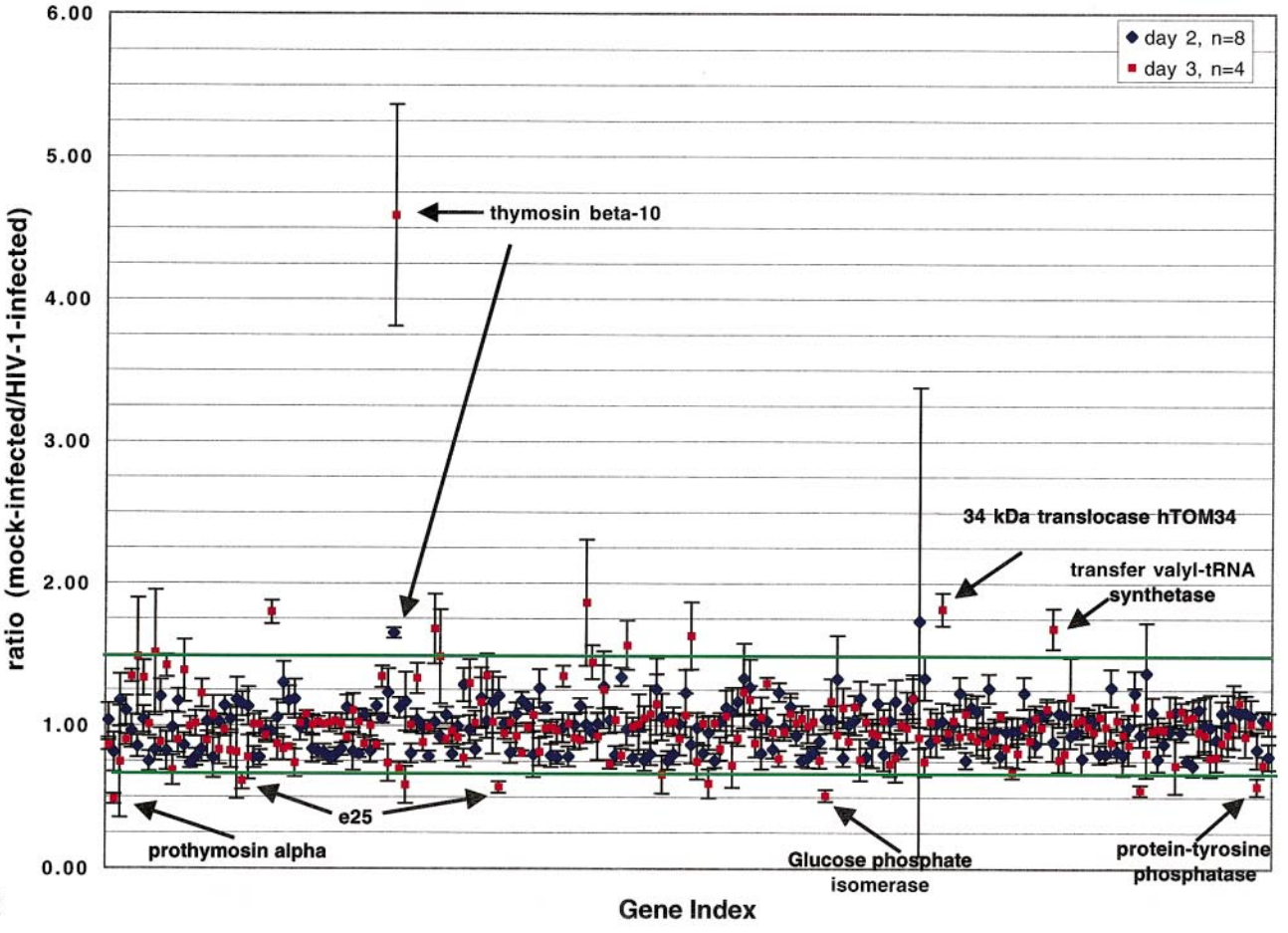
The amount of a given mRNA in each sample was determined by measuring the Cy3 and Cy5 signal of each spot after scanning the microarray at 532 and 633 nm, respectively. The raw data were collected, normalized, and statistically analyzed using the custom software as outlined in the Appendix. Using a strategy similar to that described by Chen and colleagues, differentially expressed genes were selected based on the distribution of the ratios at each time point [(*9*) and see Appendix]. The list of differentially expressed genes that are greater than 1.5-fold up- or downregulated is presented in Table 1. Those mRNAs with high relative errors in the ratio (greater than 25%) or containing repeated elements were excluded from this list. The majority of differences in gene expression were observed at day 3 p.i., with 17 of

FIG. 1. Pseudo-color image from a cDNA microarray hybridized with fluorescent probes from mock- and HIV-1-infected samples. (A) Microarray results for the complete set of 1506 genes hybridized with labeled mock- and HIV-1-infected cDNAs 3 days p.i. On the left-hand side of A, mock-infected sample was labeled in Cy3 (represented in green) and HIV-1-infected sample was labeled in Cy5 (represented in red). This labeling scheme was reversed in the right-hand side of A. (B) A section of a microarray containing 512 cDNAs hybridized with probes generated from cells 2 days p.i. Arrows indicate cDNA clones corresponding to the following genes: (1) GFP control, (2) HIV-1 pol gene, (3) ribosomal protein L21, (4) prothymosin-α, (5) TCR-α, and (6) CD8-α. The labeling scheme was exactly as described for A. (C) The same section of the microarray as shown in B, except RNA was isolated from cultures 3 days p.i. Genes expressed at equal levels in both cell populations will appear yellow, while differentially expressed genes will appear as shades of either red or green. The brightness of each spot reflects the relative abundance of that gene in the two cell populations.

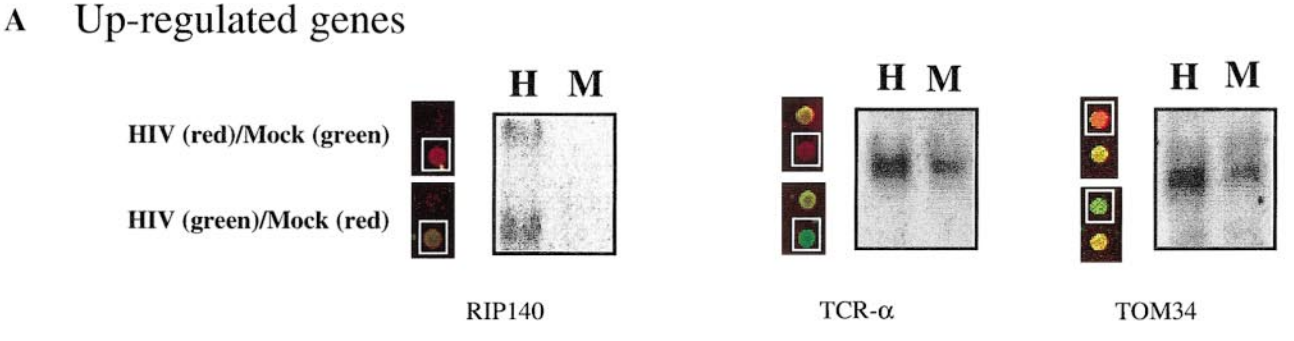
FIG. 2. Comparison of ratio values for 200 highly expressed genes. A line plot of the normalized ratio with errors (one standard deviation) for 200 of the most highly expressed cellular genes. Day 2 ratios are shown by blue diamonds, and day 3 ratios are shown by red squares. The putative identifications for several differentially expressed genes are indicated. Upregulated genes are above 1 on the *y*-axis and downregulated genes are below 1. A value of 0.5 is equal to a twofold downregulation with respect to HIV-infected cells. Genes believed to be of special interest are indicated.

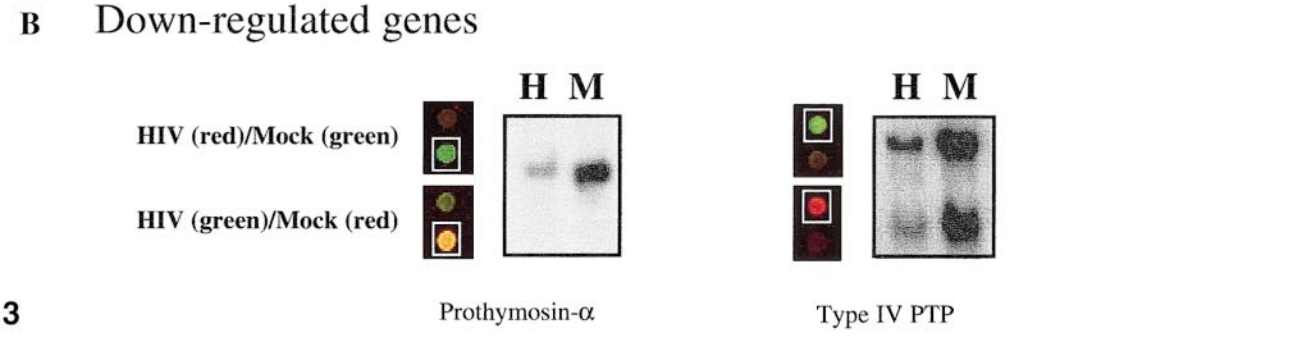
FIG. 3. Differential expression of select cellular genes. (A) Individual microarray and Northern blot results for several genes upregulated by HIV-1 infection. Twenty micrograms of total RNA from HIV-1-infected (H) or mock-infected (M) cells 3 days p.i. was hybridized with probes derived from the PCR products that were spotted on the microarray. In all cases the signal was normalized to the ribosomal L21 gene that was unchanged at both 2 and 3 days p.i. (data not shown). The labeling scheme used is indicated. (B) The same analysis was performed for several downregulated genes.


A
B
1
2
3
6
4
5
C
1
2
3
6
4
5


6.00
5.00
4.00
3.00
2.00
1.00
0.00
ratio (mock-infected/HIV-1-infected)
day 2, n=8
day 3, n=4
thymosin beta-10
34 kDa translocase hTOM34
transfer valyl-tRNA synthetase
e25
prothymosin alpha
Glucose phosphate isomerase
protein-tyrosine phosphatase
Gene Index

2


A Up-regulated genes
H M
HIV (red)/Mock (green)
HIV (green)/Mock (red)
RIP140
TCR-α
TOM34


B Down-regulated genes
H M
HIV (red)/Mock (green)
HIV (green)/Mock (red)
Prothymosin-α
Type IV PTP

3

20 genes being differentially regulated exclusively at this time point. The reproducibility and accuracy of our measurements were further supported by the fact that two independent cDNAs of the same gene, the putative integral membrane protein, e25, were downregulated to the same extent in both cases (Table 1 and Fig. 2).

Given that expression level measurements showing only 1.5-fold differences may not be considered significant in more traditional biological measurements (e.g., Northern or Western analyses), it was important to demonstrate that the 1.5-fold differences were statistically significant. The normalized expression ratios for the 200 most highly expressed genes were plotted with error bars in Fig. 2. On the same graph, horizontal lines indicate 1.5-fold expression changes in both directions. This graph clearly shows that most of the selected genes were outside the norm not only in their expression ratios but also by more than the variation of the measurements and labeling processes (indicated by error bars). Indeed, from the data it is apparent that expression ratios as low as 1.2 or 1.3 can be statistically meaningful. In addition, the microarray ratios may be underestimated when compared with more traditional approaches such as Northern blot (Fig. 3).

A set of up- and downregulated clones were selected for confirmation by Northern blots. The array data in both color schemes are shown with the corresponding Northern blots (Figs. 3A and 3B). Upregulated genes are represented as a red spot when an HIV-1-infected sample was labeled with Cy5 and green when labeled when Cy3 (compare upper and lower spots in Fig. 3A). Downregulated genes exhibited the opposite pattern when labeled the same way (compare upper and lower spots in Fig. 3B). Northern blots were performed using the cDNA clone that was spotted on the array. In six of the seven clones tested, the Northern blots were consistent with microarray results (Fig. 3 and see below for exception). The estimated sizes for mRNA species for these six genes were as follows: nuclear receptor interacting protein 140 (RIP140) (1.7 and 3.8 kb), TCR-α (1.7 kb), TOM34 (1.8 kb), prothymosin-α (1.2 kb), and type IV protein tyrosine phosphatase (PTP) (1.8 and 4.5 kb). These sizes corresponded well with published results, with the exception of RIP140. We observed two mRNA species of sizes (above) neither of which agree with the 7.5-kb transcript reported in human breast cancer line ZR75-1 (*10*). Finally, the type IV PTP, a human homologue of rat PRL-1 (*11*), also exhibited two mRNA species, both of which were downregulated by Northern blot. The largest mRNA reported in humans is approximately 1.5 kb, which may correspond to the smaller mRNA species seen in our experiments. The larger mRNA may represent an alternative polyadenylation signal, as reported for the rat gene or one of several PRL-1 homologues (*12, 13*). In general, changes determined by Northern blot were greater than those determined by microarray. The changes by Northern analysis (normalized to ribosomal protein L21) for the selected genes were as follows: RIP140 (>5 upper and lower), TCR-α (4.4), TOM34 (2.2), prothymosin-α (7), and type IV PTP (2.9 upper, 1.8 lower). Given these results and the reproducibility of the microarray experiments, genes with ratios as low as 1.3- or 1.4-fold are likely to be worth further investigation.

Of the genes tested, only thymosin-β_{10} could not be verified by Northern analysis. This result was somewhat unexpected because this gene seemed to be consistently and significantly regulated at both 2 and 3 days p.i. (Table 1 and Fig. 2). Control experiments (Northern blots and microarray hybridizations with the full length HIV-1 genome, not shown) indicated that this gene cross-hybridized with viral sequences, resulting in a false-positive signal. No other upregulated cellular gene listed in Table 1 was found to cross-hybridize to viral sequences. These results indicate that cross-hybridization could be a potential problem in experimental systems in which abundant exogenous RNA species are present (e.g., viral or bacterial infections) and appropriate confirmatory steps must be taken to avoid false positives.

Discussion. We have examined the expression of 1506 cDNAs during HIV-1 infection of CEM cells using cDNA microarrays and identified 20 differentially expressed genes that function in a variety of cellular pathways. These results support the hypothesis that HIV-1 infection alters cellular gene expression at a global level that has not been recognized previously. To our knowledge, this study is the first quantitative large-scale analysis of host cell gene expression in HIV-1-infected cells and further demonstrates the usefulness of microarray technology in the study of pathogenic viruses. In addition, we have developed methods to quantitate and analyze microarray images that can be used in our subsequent studies of viral–host cell interactions. Although published data on reproducibility and error analysis are limited, we believe the analysis methods used for this system will be transferable to other array systems and technologies. Regardless, users are cautioned to perform appropriate statistical analyses before pursuing leads based on apparent differential expression ratios because of a large number of false-positive results from lowly expressed genes with high, effectively meaningless ratios.

The differentially expressed genes identified in this study have been reported to function in a variety of cellular processes including T-cell receptor-mediated signaling, subcellular trafficking, transcriptional regulation, and a variety of cellular metabolic pathways. In addition, several ESTs and as yet uncharacterized mRNAs were influenced by HIV-1 infection. Presumably, some of these genes will play a role in the host cell defense mechanisms and/or facilitation of the viral life cycle. Several genes are worth commenting on, given their association with HIV or other viral infections or their

known roles in cellular regulatory pathways that have the potential to be influenced by viral infection.

T-cell Interactions. Several of the differentially expressed genes are known to participate in T-cell interactions. The mRNAs for the T-cell receptor-α chain and CD8-α, both involved in recognition of antigen via interaction with MHC molecules, were upregulated at day 3 p.i. increasing from barely detectable at day 2 p.i. to differentially expressed at 3 days (compare arrows 5 and 6, Figs. 1B and 1C). Although the functional significance of this observation is not yet known, it could partially account for populations of HIV-1-infected CD8+ cells in infected individuals (*14*) or the increase in CD8+ cell counts observed in some patients (*15*). The mechanisms of CD8 regulation under these experimental systems are not known and could be mediated directly by HIV-1 proteins or indirectly through modulation of cytokine production by HIV-1-infected cells. Other interesting differentially expressed genes (in the context of T-cell signaling) include the Txk kinase (*16*), a member of the Tec family of protein kinases, and the recently identified type IVA PTP (*17*), similar to rat PRL-1 (*13*).

Cellular Trafficking/Transport. Translocase of the outer membrane (TOM34), a newly identified member of the mitochondrial transport machinery (*18*), was found to be upregulated 1.8-fold at day 3. In addition to its role in mitochondrial import, TOM34 is a putative member of the tetratricopeptide repeat (TPR) family of proteins (*19*). Given that the HIV-1 proteins Vpu and Gag have been shown to interact with a cellular TPR-containing Vpu binding protein (UBP) (*20*), it may be interesting to examine TOM34 for similar interactions.

Another cellular gene regulated by HIV-1 with a potential role in cellular trafficking (*21*) is Prothymosin-α. Interestingly, this evolutionary conserved and widely distributed acidic protein has been shown to interact *in vitro* with the leucine motif activation domain of HIV-1 Rev (*22*), a viral protein involved in nuclear export of unspliced viral RNAs [for review see (*3*)]. Since both rev and prothymosin-α may be involved in nuclear export and/or shuttling, this physical interaction could be of functional importance. Therefore, the downregulation of this gene in the context of HIV-1 infection is worth further investigation. For instance, will high levels of prothymosin-α inhibit HIV-1 replication, possibly by sequestering and inactivating the viral rev protein?

Transcriptional Regulation. The transcriptional regulator RIP140 (*10*) was found to be upregulated at 3 days post-HIV-1 infection. RIP140 is now considered a cofactor for several members of the nuclear receptor superfamily and is a negative regulator of peroxisome proliferator-activated receptor (PPAR)-mediated transcription (*23*). Since PPAR controls expression of several genes involved in lipid metabolism, these results could suggest that lipid metabolism is perturbed in HIV-1-infected cells. Given the problem of dorsocervical fat pad enlargement (buffalo hump, especially in HIV-infected patients that have not been exposed to protease inhibitor therapy) (*24*), this should be vigorously investigated. Alternatively, RIP140 may be recruited to participate in the regulation of transcription from the viral LTR, since several groups have reported nuclear receptor binding sites in the HIV-1 LTR, including PPAR (*25*). Of course, additional studies will be required to obtain a more complete global perspective on the effects of HIV-1 infection on host cell gene expression. However, by examining a variety of HIV-1 infection systems, including different viral strains or cell types, more complete time courses, and larger sets of human genes, certain pathways will emerge as playing critical roles during viral infections.

Materials and Methods. HIV-LAI viral stock, CEM cells, and the conditions for infection were as described previously (*8*). Cell viability was determined by trypan blue dye exclusion and percentage infection confirmed by immunofluorescence assay (IFA) [(*8*) and data not shown].

Human cDNA IMAGE clones were purchased from Research Genetics (Huntsville, AL; UG Build No. 18 V5.0, Plate Nos. 501–505, 511–514, and 516–520 R2). cDNA inserts for IMAGE clones and controls were PCR amplified, precipitated, and purified.[2] DNA pellets were suspended in a 50% solution of Reagent D (proprietary solution, Amersham–Pharmacia Biotech, Buckinghamshire, England) and deposited on 75 × 25-mm coated glass microscope slides (Amersham Type 4, Type VP) with the use of a Molecular Dynamics (Sunnyvale, CA) "Generation II" Microarray Spotter II. After spotting, microarrays were air-dried, cross-linked at 450 mJ, and stored desiccated under N_2 until needed.

The GFP constructs used for construction of target cDNA and *in vitro* synthesized GFP poly(A)$^+$ RNA were kindly provided by Dr. Pete Nelson (Department of Molecular Biotechnology, University of Washington). GFP cDNA was amplified by PCR and deposited at predetermined positions on the microarray. GFP poly(A)$^+$ RNA was generated by *in vitro* transcription of 0.5 μg of template DNA (plasmid pSP64-GFP linearized with *Eco*RI) using SP6 RNA polymerase (Gibco BRL, Gaithersburg, MD) as described in the manufacturer's protocol. *In vitro* transcribed GFP poly(A)$^+$ was added to both RNA samples as a control prior to reverse transcription (see below).

Total RNA was extracted and Northern blots were carried out as described previously (*8*). Total RNA was treated with RNase-free DNAse I and extracted with phenol/chloroform, and poly(A)$^+$ mRNA was isolated by a single pass over an oligo(dT) cellulose column (Collab-

[2] Detailed protocols on PCR amplification and purification and a complete list of control and IMAGE cDNAs are available upon request.

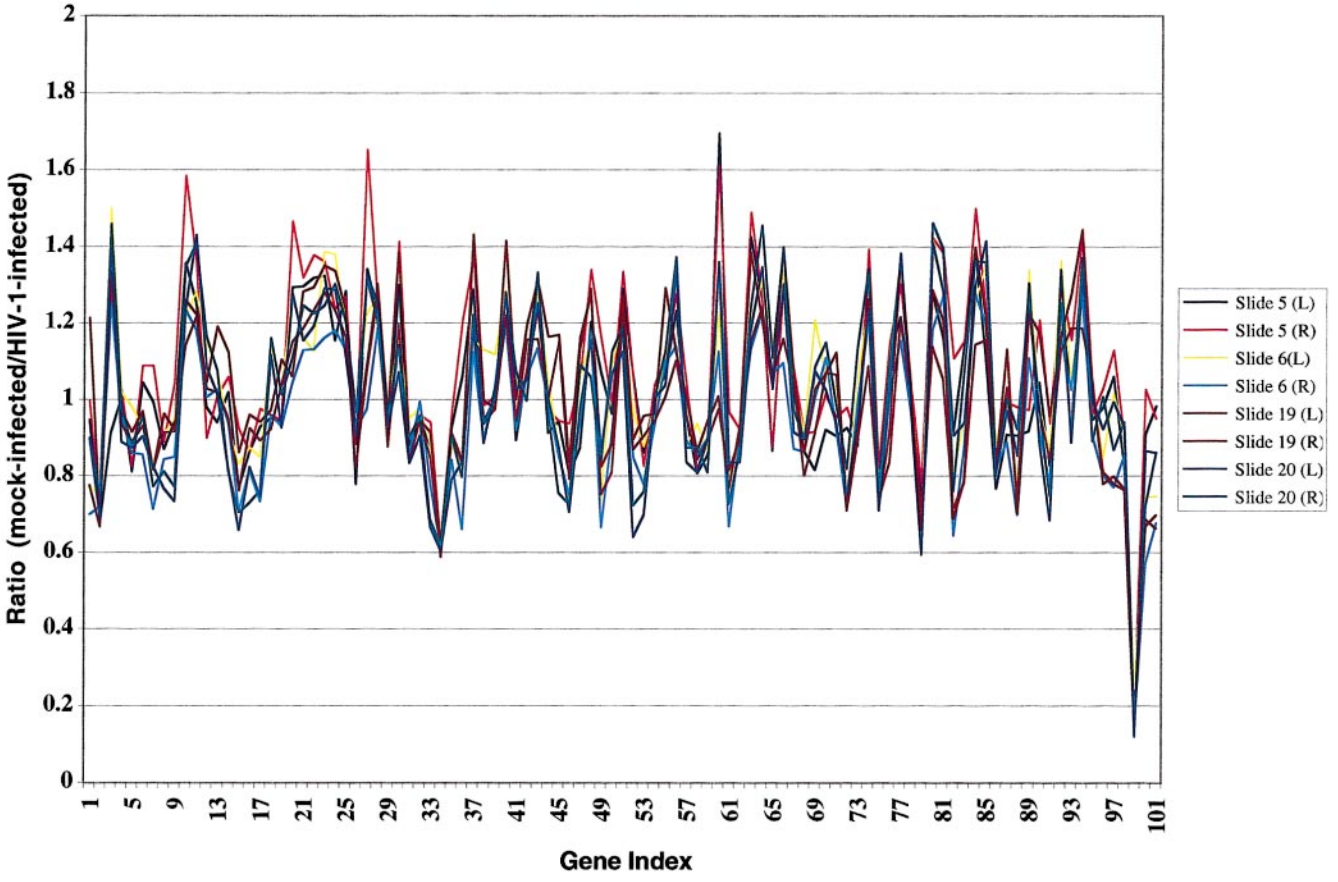


FIG. A1. Reproducibility of the ratio measurement for the 100 most abundant cDNAs. The ratios for duplicate spots from four separate microarrays were normalized as described in the text. The normalized ratio values (*y*-axis) for the 100 most abundant cDNAs (*x*-axis) 2 days p.i. are shown superimposed on the line plot. Data points were taken from duplicate spots from four separate microarrays hybridized with probes using both labeling schemes. Each data set is indicated by a different color line.

orative Biomedical Products). Fluorescent-labeled cDNA probes were generated by reverse transcription (RT) as follows: 1 μg of poly(A)$^+$ RNA from mock- and HIV-infected cells, 5 ng of GFP poly(A)$^+$ RNA, and 8 pmol of anchored dT primer (Amersham) were combined in a 10.5-μl reaction volume. The solution was heated to 70°C for 10 min, chilled briefly on ice, and centrifuged. Reverse transcription was performed in a 20-μl reaction volume. Final concentrations were 1× First Strand Buffer (Gibco BRL), 10 mM DTT, 50 nM dATP/dGTP/dTTP, 6.25 nM nonlabeled dCTP, 6.25 nM FluoroLink-dCTP (either Cy3 or Cy5 labeled, Amersham), 0.5 units/μl placental RNase inhibitor (Amersham), and 10 units/μl Superscript II RT (Gibco, BRL). The contents were mixed and incubated at 42°C for 2 h. RNA was removed by heating the sample to 94°C for 3 min followed by treatment with sodium hydroxide (0.25 N final) for 10 min at 37°C. The sample was neutralized by addition of an equal amount of HCl, followed by 1 M Tris (pH 7.5) to 0.2 M. Unincorporated nucleotides were removed using 96-well multiscreen FB filter plates (Millipore).

Prior to hybridization, microarrays were rinsed briefly in deionized H_2O and dried with spectroscopy-grade isopropanol to remove residual dust. Fluorescently labeled first-strand cDNAs were concentrated by drying and resuspended in 20 μl of 1× hybridization solution [5× SSC, 5× Denhardt's solution, 1% SDS, 50% formamide, 0.1 μg/μl denatured salmon testes DNA, and 20 μg/ml poly(A)$_{72}$]. The appropriate Cy3- and Cy5-labeled probes were combined, denatured by boiling, and applied to the microarray under a 22 × 64-mm glass coverslip. Microarrays were incubated at 42°C in a custom-built humidified chamber for 16–20 h. Microarrays were washed once in 2× SSC/0.1% SDS at 37°C for 5 min, once in 0.2× SSC/0.1% SDS at room temperature for 5 min, twice in 0.1× SSC at room temperature for 2 min each, and once with deionized H_2O for 10 s. Slides were then dried under nitrogen and scanned with a confocal dual-laser scanner at 532 and 633 nm (Molecular Dynamics).

APPENDIX

Image Processing, Data Normalization, and Error Analysis. An important issue for successful use of microarrays in examining biological problems is data analysis. As pointed out previously (*9, 26*), a focus of the data

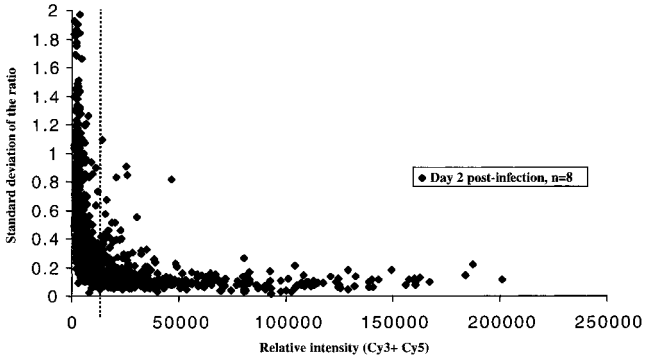


FIG. A2. Error analysis for all 1536 cDNAs as a function of overall signal intensity. The standard deviation of the ratio for each clone (*y*-axis) was plotted against the sum of intensity for the Cy3 and Cy5 channels (*x*-axis), $n = 8$. Data represent the cumulative results for experiments on day 2 p.i., $n = 8$. The minimal threshold value of 13,000 relative intensity units is indicated by the dashed line.

analysis should be to generate both expression ratios and estimates of uncertainties in these ratios. We have developed custom array analysis software (E. Hammersmark and R. Bumgarner, in preparation) in order to provide estimates of error in the ratios for each independent analysis. Briefly, the software finds the spots in a composite image constructed from the normalized sum of both color channels, extracts the intensity data from both color channels by integrating in a circle around each spot center, determines the local background and variance in local background by measuring intensity in several locations around each spot, and estimates the error in the ratio by determining the ratio in several regions of the spot and calculating a standard deviation. Each image is analyzed independently to produce expression ratios and error estimates in these ratios. This allows us to discard data of low confidence based on a single spot on the array and to investigate the reproducibility of the arrays across multiple repeated measurements.

The image analysis produces nonnormalized expression ratios, e.g., the output is the intensity of (Cy3 signal minus background) divided by (Cy5 signal minus background), and is uncorrected for labeling or fluorescence detection efficiencies and for relative mRNA concentrations. It is common practice to normalize data of this type by dividing by a linear scaling factor, such as a set of "housekeeping" genes, so that these expression ratios are set to 1. From our data (day 2 p.i., $n = 8$) it was apparent that the Cy3 signal did not increase with concentration at the same rate as the Cy5 signal under normal hybridization conditions (data not shown). That is, a linear normalization based on the housekeeping genes overcorrected the Cy3 signal at lower signal levels. Hence, we developed an algorithm to normalize the data with a factor that varied as a function of intensity (R. E. Bumgarner and E. Hammersmark, in preparation). The net result of this normalization procedure is that the eight measurements of the data sets on day 2 p.i. agree within 10% relative error for the 100 most highly expressed genes and within 15% relative error for the 500 most highly expressed genes (Fig. A1 and data not shown). The data obtained from day 3 p.i. were treated in the same manner.

The data sets for experiments on days 2 and 3 p.i. were normalized and combined to produce an average expression ratio and standard deviation for both data sets. The standard deviation was plotted as a function of expression level (defined as the sum of the signal in both color channels) and used to set a threshold intensity below which we have little confidence in the expression ratios (Fig. A2). An intensity cutoff was chosen such that the relative error in the ratios was less than 25%. This intensity cutoff also corresponded well with the lower limit of what we could confirm by visual reinspection of the array

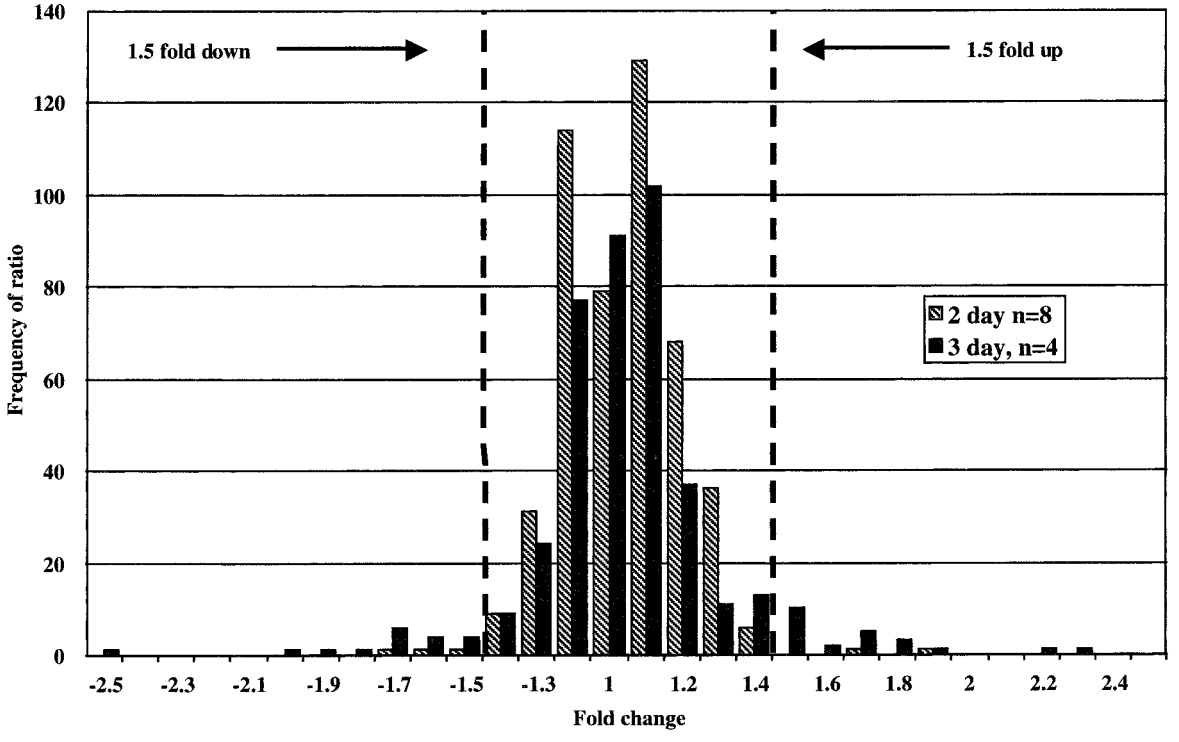


FIG. A3. Distribution of ratio values 2 and 3 days p.i. A histogram represents the distribution of ratios for genes above the minimum threshold intensity. Dashed lines indicate the cut-off ratio of 1.5-fold, corresponding to approximately the 5th and 95th percentiles (the top 5% of up- and downregulated genes). Day 2 ratios are shown by hatched bars; day 3 ratios are displayed by solid gray bars. For the purpose of this graph, downregulated ratios are shown as the negative inverse values of the normalized ratio.

images. In our data set, 477 cDNAs at day 2 p.i. and 436 cDNAs at day 3 p.i. exhibited hybridization signals above their respective minimum threshold intensities at each time point. The remainder of the cDNAs (roughly two-thirds of the spotted clones) did not produce data with sufficient signal-to-noise ratios to be analyzed accurately. Our study examined a total of about 1500 genes, of which ~30% produced acceptable signals above threshold.

To assess the validity of our analysis, we examined the normalized ratios for sets of control genes. As expected, the average ratios for the GFP control gene and a set of 15 housekeeping genes were close to 1; for example, see arrow 1 (GFP) and arrow 3 (ribosomal protein L21) in Fig. 1. Conversely, the HIV-1 cDNAs, used as a positive control for upregulated genes, showed average changes between 13- and 32-fold for days 2 and 3 p.i., respectively (Fig. 1, arrow 2).

Selection and Identification of Differentially Expressed Genes. The frequencies of the HIV-1/mock expression ratios for cellular genes on days 2 and 3 p.i. are plotted in a histogram in Fig. A3. Differentially expressed genes that lie in the upper and lower 4% of the histogram, i.e., the genes that were the most differentially expressed, were selected. This corresponded to ratios that were approximately 1.5-fold up- or downregulated. It is clear from this histogram that gene expression remains relatively constant at the day 2 time point, with most genes clustered around 1.2-fold in either direction. In contrast, the distribution of ratios at day 3 is more varied, with multiple outliers exhibiting ratios greater than 1.5-fold.

ACKNOWLEDGMENTS

We gratefully acknowledge the members of Dr. Leroy Hood's laboratory, especially Drs. Michèl Schummer and Pete Nelson, for their expert advice and patience during the initial phases of this project. In addition, we are grateful to our collaborators Dr. Leroy Hood and Molecular Dynamics for providing access to the microarrayer and scanner. We also thank members of our weekly array group for critical discussions and direction, Dr. Marcus Korth, Matthew T. Czislowski, and Matthew D. Taylor. G. K Geiss is supported in part by NIH Training Grant AI07I40 via the University of Washington Center for AIDS Research Program. Other support was provided by NIH Research Grants RR00166 and AI41535. Mahru An is a recipient of a Howard Hughes Medical Institute Research Internship.

REFERENCES

1. Baca, L. M., Genis, P., Kalvakolanu, D., Sen, G., Meltzer, M. S., Zhou, A., Silverman, R., and Gendelman, H. E. (1994). Regulation of interferon-α-inducible cellular genes in human immunodeficiency virus-infected monocytes. *J. Leuk. Biol.* **55,** 299–309.
2. Thornton, A. M., Buller, R. M., DeVico, A. L., Wang, I. M., and Ozato, K. (1996). Inhibition of human immunodeficiency virus type 1 and vaccinia virus infection by a dominant negative factor of the interferon regulatory factor family expressed in monocytic cells. *Proc. Natl. Acad. Sci. USA* **93,** 383–387.
3. Emerman, M., and Malim, M. H. (1998). HIV-1 regulatory/accessory genes: Keys to unraveling viral and host cell biology. *Science* **280,** 1880–1884.
4. Brown, P. O., and Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* **21,** 33–37.
5. Lipshutz, R. J., Fodor, S. P., Gingeras, T. R., and Lockhart, D. J. (1999). High density synthetic oligonucleotide arrays. *Nat. Genet.* **21,** 20–24.
6. Lennon, G. G., and Lehrach, H. (1991). Hybridization analyses of arrayed cDNA libraries. *Trends Genet.* **7,** 314–317.
7. Zhu, H., Cong, J.-P., Mamtora, G., Gingeras, T., and Shenk, T. (1998). Cellular gene expression altered by human cytomegalovirus: Global monitoring with oligonucleotide arrays. *Proc. Natl. Acad. Sci. USA* **95,** 14470–14475.
8. Agy, M. B., Wambach, M., Foy, K., and Katze, M. G. (1990). Expression of cellular genes in CD4 positive lymphoid cells infected by the human immunodeficiency virus, HIV-1: Evidence for a host protein synthesis shut-off induced by cellular mRNA degradation. *Virology* **177,** 251–258.
9. Chen, Y., Dougherty, E. R., and Bittner, M. L. (1997). Ratio-Based decisions and the quantitative analysis of cDNA microarray images. *J. Biomed. Optics* **2,** 364–374.
10. Cavailles, V., Dauvois, S., L'Horset, F., Lopez, G., Hoare, S., Kushner, P. J., and Parker, M. G. (1995). Nuclear factor RIP140 modulates transcriptional activation by the estrogen receptor. *EMBO J.* **14,** 3741–3751.
11. Diamond, R. H., Cressman, D. E., Laz, T. M., Abrams, C. S., and Taub, R. (1994). PRL-1, a unique nuclear protein tyrosine phosphatase, affects cell growth. *Mol. Cell. Biol.* **14,** 3752–3762.
12. Zeng, Q., Hong, W., and Tan, Y. H. (1998). Mouse PRL-2 and PRL-3, two potentially prenylated protein tyrosine phosphatases homologous to PRL-1. *Biochem. Biophys. Res. Commun.* **244,** 421–427.
13. Carter, D. A. (1998). Expression of a novel rat protein tyrosine phosphatase gene. *Biochim. Biophys. Acta* **1442,** 405–408.
14. Kitchen, S. G., Uittenbogaart, C. H., and Zack, J. A. (1997). Mechanism of human immunodeficiency virus type 1 localization in CD4-negative thymocytes: Differentiation from a CD4-positive precursor allows productive infection. *J. Virol.* **71,** 5713–5722.
15. Margolick, J. B., and Donnenberg, A. D. (1997). T-cell homeostasis in HIV-1 infection. *Semin. Immunol.* **9,** 381–388.
16. Hu, Q., Davidson, D., Schwartzberg, P. L., Macchiarini, F., Lenardo, M. J., Bluestone, J. A., and Matis, L. A. (1995). Identification of Rlk, a novel protein tyrosine kinase with predominant expression in the T cell lineage. *J. Biol. Chem.* **270,** 1928–1934.
17. Montagna, M., Serova, O., Sylla, B. S., Feunteun, J., and Lenoir, G. M. (1995). A 100-kb physical and transcriptional map around the EDH17B2 gene: Identification of three novel genes and a pseudogene of a human homologue of the rat PRL-1 tyrosine phosphatase. *Hum. Genet.* **96,** 532–538.
18. Nuttall, S. D., Hanson, B. J., Mori, M., and Hoogenraad, N. J. (1997). hTom34: A novel translocase for the import of proteins into human mitochondria. *DNA Cell Biol.* **16,** 1067–1074.
19. Young, J. C., Obermann, W. M. J., and Hartl, F. U. (1998). Specific binding of tetratricopeptide repeat proteins to the C-terminal 12-kDa domain of hsp 90. *J. Biol. Chem.* **273,** 18007–18010.
20. Callahan, M. A., Handley, M. A., Lee, Y. H., Talbot, K. J., Harper, J. W., and Panganiban, A. T. (1998). Functional interaction of human immunodeficiency virus type 1 Vpu and Gag with a novel member of the tetratricopeptide repeat protein family. *J. Virol.* **72,** 5189–5197. [published erratum appears in *J. Virol.* **72**(10), 8461, 1998]
21. Manrow, R. E., Sburlati, A. R., Hanover, J. A., and Berger, S. L. (1991). Nuclear targeting of prothymosin alpha. *J. Biol. Chem.* **266,** 3916–3924.
22. Kubota, S., Adachi, Y., Copeland, T. D., and Oroszaln, S. (1995). Binding of human prothymosin-a to the leucine-motif/activation domains of HTLV-1 Rex and HIV-1 Rev. *Eur. J. Biochem.* **233,** 48–54.
23. Miyata, K. S., McCaw, S. E., Meertens, L. M., Patel, H. V., Rachubinski, R. A., and Capone, J. P. (1998). Receptor-interacting protein 140 interacts with and inhibits transactivation by peroxisome proliferator-activated receptor alpha and liver-X-receptor alpha. *Mol. Cell. Endocrinol.* **146,** 69–76.
24. Lo, J. C., Mulligan, K., Tai, V. W., Algren, H., and Schambelan, M. (1998). "Buffalo hump" in men with HIV-1 infection. *Lancet* **351,** 867–870.
25. Ladias, J. A. (1994). Convergence of multiple nuclear receptor signaling pathways onto the long terminal repeat of human immunodeficiency virus-1. *J. Biol. Chem.* **269,** 5944–5951.
26. Ermolaeva, O., Rastogi, M., Pruitt, K. D., Schuler, G. D., Bittner, M. L., Chen, Y., Simon, R., Meltzer, P., Trent, J. M., and Boguski, M. S. (1998). Data management and analysis for gene expression arrays. *Nat. Genet.* **20,** 19–23.